

DOCUMENT RESUME

ED 128 470

TM 005 667

AUTHOR Steinmetz, Andres
TITLE Program Evaluating vs. Program Improvement and Some Implications for Training Evaluators.
PUB DATE [Apr 76]
NOTE 33p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)
EDRS PRICE MF-\$0.83 HC-\$2.06 Plus Postage.
DESCRIPTORS Educational Objectives; Evaluation Criteria; Evaluation Methods; Formative Evaluation; *Professional Personnel; Professional Training; *Program Evaluation; *Program Improvement; Research Design; Social Influences; *Standards

ABSTRACT

To evaluate a specific program means to compare it against a standard that specifies what the program should be like at a specific time. Standards may be constructed in three ways: surface standards force the definition into a model shaped by scientific procedure; deep standards include information on the inputs, processes, and outputs for each component and subcomponent; and profound standards reach deeper into the organization than its task structure and exhaustively covers all dimensions of organizational functioning, and are set by the program staff. In this situation then, the role of the evaluator includes expressing and explicating the standards set by the program staff, and confronting management with the decisions they must make. The evaluator must also look at the broader environment of the educational organization, i.e., the sociotechnical systems of which schools are a part. This broader perspective implies that the ability to build models, and the ability to be able to gather data relative to a large variety, of different phenomena, and to report this data using print, verbal, nonverbal, visual and auditory media be included as part of an evaluator's training. (Author/BW)

* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *

ED128470

TM

PROGRAM EVALUATION vs PROGRAM IMPROVEMENT
AND SOME IMPLICATIONS FOR
TRAINING EVALUATORS

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
PERSON OR ORGANIZATION ORIGIN-
AL. POINTS OF VIEW OR OPINIONS
DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY.

Andres Steinmetz
Evaluation Research Center
University of Virginia
164 Rugby Road
Charlottesville, Virginia

A paper presented at the meeting of the
American Educational Research Association,
April 21, 1976
San Francisco, California

TM005 606

Table of Contents

	<u>Page</u>
A. Definition.....	1
B. Imposing Standards.....	3
C. Discerning Standards.....	4
D. Building Standards	
1. Surface standards.....	5
2. Deep standards.....	6
3. Profound standards.....	14
a. The role of the Program Improvement Specialist (PIC).....	15
b. Sociotechnical systems.....	18
c. Implications for training PICs.....	23
d. Generic subsystem categories.....	26
e. Parson's theory of action.....	28

A. Definition.

We can say that whenever we do evaluation we get involved in making comparisons. We compare the object we are interested in evaluating against something else which we call a Standard. The Standard is made up of qualities, characteristics, behaviors, conditions, etc., which are considered desirable and which the object to be evaluated is expected to manifest. Judgments about the worth of that object will then be based on the extent to which one is able to determine that the object meets the characteristics of the Standard. Another way of saying this is to say that in order to evaluate an object we must have available a Standard, S. Performance measures, P, taken on the object, are aimed at determining the existence of the specifications contained in S. The discrepancies, D, found between S and P across all the variables examined then become the basis for a judgment about the worth of the object.

This process of setting S and determining P and D seems the same no matter what is being evaluated, although it may be much more difficult to carry out in one case than in another. For example, suppose one wants to evaluate a plate lunch. The evaluation would require or imply a Standard which specifies the dimensions, qualities, characteristics and so on which are going to be taken as desirable. In that sense one gets involved in constructing an ideal plate lunch or a model of one. Once a description is available of what a plate lunch should be like, one may proceed to look at a specific one to see the extent to which it meets the

specifications called for by the Standard. Now it may be in this case that some of the qualities called for by the Standard having to do for example with taste or smell may be hard to describe in an observable form. It may be hard to derive indices of these qualities and hard to make valid and reliable comparisons between a specific plate lunch and the ideal. That difficulty, however, does not detract from the fact that P is being compared to S and that discrepancy information is being generated. It is easy to see that if we were evaluating a table, we could create a very observable model of a table which we can imbue with most of the characteristics we desire and thus refer to as our Standard. Evaluating a particular table would then involve taking Performance measures as called for or implied by the Standard (for example, concerning weight, size, type of wood, form, etc.) and determining the discrepancy between the actual state of affairs and the desired state of affairs.

Thus, to evaluate a specific program would mean to compare it against some Standard. One would be making that comparison at a particular time and since a program is an ongoing activity, the comparison would be in respect to the state or condition of the program at that time. A proper Standard would therefore specify what the program should be like at that time. But one may also speak of a program as a collection of related and interdependent activities organized to fulfill some purpose and thus Standards may have to exist or be constructed for many different aspects or dimensions of a program in the course of its evaluation.

B. Imposing Standards.

It should be evident that, according to the view of evaluation presented above, no evaluation is possible without a Standard. However, more often than not, when attempting an evaluation of a program, one finds its objectives, activities and organization so imbedded in unarticulated experience and so undifferentiated that Standards are hard to identify. One common way of dealing with such a situation is to outline the requirements of scientific investigation and point out the specific ways in which program operations are ambiguous and thus fail to provide a Standard. The adequacy of a Standard is thus related to the ease with which certain canons of scientific procedure can be applied in the collection of Performance information. In other words, the implication is that the program has to run in a certain manner and that its activities, objectives and plans have to be portrayed in a certain manner in order for the program to be judged adequate up to data collection; that is, judged to meet certain specifications which stand as prerequisites to the application of scientific procedure. This way of proceeding amounts to imposing as a Standard the rules governing a particular way of doing evaluation.

Another way in which Standards are often imported and laid onto a program is to subject it to examination by an expert. The expert would be somebody with extensive experience in the problems and concerns faced by the program. This person will have formed some ideas about better and worse ways of doing whatever the program is doing and will judge the program accordingly. Exactly what Standards are invoked, however, may often not be made fully explicit nor will, necessarily, all observations

about the program that go into making the judgments. The expert's experience is assumed as the Standard and the evaluation consists of judgments concerning the extent to which empirical and impressionistic performance observations match that Standard--the possibility that the program never completely intended to approximate the Standards represented by the experience of the expert notwithstanding.

C. Discerning Standards.

An alternative which avoids comparing a program against a set of Standards identified with expertness, and which avoids making the application of certain rules of evaluation procedure a prerequisite to evaluation, is to engage in an essentially descriptive study aimed at articulating the Standards that govern program activity whatever they are. This would entail doing a quasi-ethnographic study, exhibiting more and less recognized Standards, and contrasting these with empirical evidence relative to program operation. This approach would have the merit of making the program understandable in its context. It would describe and contrast Standards and Performance and it would do so relative to the here and now situation and relative to external contingencies and the personalities of the individuals involved. It would thus render the program understandable by relating it to the life of the actors involved--in fact by presenting it as a living enterprise. However, a detailed representation of the culture of a program which does not exhibit the idealized characteristics of scientific action is often automatically perceived as being somehow deficient even by program staff themselves. This may be because of the commonly held

notion that program activity should show the qualities of infallibility, predictability, and preciseness usually associated with scientific procedure.

D. Building Standards.

1. Surface standards. Instead of critiquing a program for its lack of adherence to the conditions necessary to scientific procedure, the evaluator may undertake action to insure that at least certain aspects of program operation are described and conducted in such a manner as to make them compatible to the rules of procedure and evidence which govern scientific investigation. This would amount to building a more or less articulated value system (that inherent to scientific procedure) into at least some aspects of the program. Most evaluations can probably be classified under this approach. For example, many evaluators proceed by working with program staff in order to define some terminal program outcomes which represent the primary purposes of the program. This will mean that the worth of the program will be a function of the extent to which these terminal outcomes are reached and therefore they represent the Standard against which the program will be judged. Evaluators and program staff will work together in order to generate operational definitions of all terms used, along with measurable or observable Performance indices. Further, since program objectives are a function of time, deadlines are set by which the indices agreed to should be observable. Data collection strategies and instruments are then developed in accordance with the usual rules of scientific procedure. Precautions are taken through the application of quasi-experimental designs to insure that program results can be attributed to program activities.

While the objectives per se may not be evaluated under this approach (that is, the objectives may not be examined for their intrinsic worth), the Standards employed are at least clear and certainly related to the intentions of the program. However, it is also true that building the Standard against which the program will be compared around the terminal objectives of the program fails to incorporate into the Standard the full range of meaning that makes the program what it is in the experience of its members. This can lead to the feeling, at least on the part of program staff, that the Standard against which the program is judged does not take into account sufficient detail about the program and thus judges it too superficially. There may be, in addition, the feeling that the operational definitions agreed to do not convey enough of the context and texture of the program and are thus simplistic and uni-dimensional. The charge is also often made that the operational definitions finally agreed to are disproportionately determined by the ready existence of acceptable measurement tools.

2. Deep standards. Some of the shortcomings of approaches to program evaluation as perceived by program staff have already been alluded to. For example, whenever Standards are not made explicit or at least agreed to by program staff, evaluators and others interested in the evaluation it is highly likely that program staff will feel their program unfairly judged. They are also likely to feel it narrowly judged when the judgment is based on stylized measurements made on certain dimensions of the program specified in advance and in accordance with certain rules of

expression. Moreover, a judgment about a program is invariably perceived as a judgment on program staff. Given such reservations about the adequacy of evaluations, it is easy to see that program staff will tend to want to protect themselves and thus the evaluator/management interaction is easily turned into an adversary relationship. Program staff naturally feel themselves accountable to funding agents, superiors in an organization, program clients, and so on, and readily think of an evaluation as a determination of the extent to which they can be said to have met their obligations. This can lead program staff to suppose that an "evaluation" will place their status, image of themselves, career plans, etc., into jeopardy. Thus, it is common to have all transactions by program staff relating to an evaluation be permeated with defense against anxiety.

A further discouragement is often suffered by program staff when evaluations are not practical or not immediately useful to the program. Evaluations tend not to be useful to program staff when they are geared to serve the interests of outsiders who associate one or two major purposes with the program and are anxious for relatively definitive and easy to understand answers concerning the extent to which those purposes are achieved. From the point of view of program staff, program evaluations tend to be simplistic when they do not take into account the actual day-to-day problems faced by a program and the often unarticulated design changes a program undergoes in an effort at remaining responsive to changing conditions and unforeseen events. Often, since important decisions about the livelihood of a program are made on the basis of just such (simplistic) evaluations, resistance to evaluation

becomes expressed in an unwillingness to submit to the requirements of experimental design since these symbolize the evaluation process. Quasi-ethnographic approaches to evaluation also carry severe restrictions. They tend to provide such a wealth of data understandable in so many different ways that their immediate usefulness to program operation is unclear. While they carry exceptional anthropological interest and are an essential basis for a comprehensive critique of educational methods, they are by nature not intended to be confrontive to the individuals involved and thus, without drastic amendment, are not a useful input to day-to-day decision making.

In order for Performance data to be immediately useful to program staff, it must be collected relative to a broader range of program dimensions than are represented by terminal outcomes. In other words, the evaluation must look deeper into the program. A number of different approaches to evaluation have been formalized in the last half dozen years which aim to provide data useful to program staff and to serve the staff's interest at improving their program.¹ For example, figure 1 shows the kinds of evaluation that may be performed around two kinds of management action: program planning and program implementation.

"Design evaluation" in figure 1 refers to judging the adequacy of program intentions. The object being evaluated here is the program plan. Any program is bound to have a basis in social and moral philosophy,

1. The remaining discussion in this section is based on the Discrepancy Evaluation Model, see A. Steinmetz, "The Discrepancy Evaluation Model" in Measurement in Education, an NCME publication, to be published in 1976.

Figure 1

KINDS OF EVALUATION USEFUL TO
PROGRAM IMPROVEMENT

Management Action	Kinds of Evaluation	Standards relative to:
Program Planning	Design evaluation	-design adequacy; construct validity, comprehensiveness, internal or logical consistency, relationship to need, appropriateness.
Program Implementation	Input evaluation	-availability of resources; extent to which prerequisites and preconditions are met, extent to which program resources (\$, people, materials, etc.) are available and deployed as planned.
	Process evaluation	-extent to which activities are carried out as planned; existence or frequency, intensity, manner, and other qualities and characteristics of the activities.
	Outcome evaluation	-interim outcomes; outcomes which are a prerequisite to other outcomes; terminal outcomes.

as well as empirical research and the critique of these bases may be referred to as the problem of construct validity. But the program plans may also be examined for their comprehensiveness, their appropriateness to the situation, their relationship to known interests and needs, and so on. An analysis may also be made to see whether the kind and level of resources being planned, the kinds and qualifications of personnel, or the kinds of materials

that appear in the plan seem adequate to support the activities that the program intends to undertake. Similarly, one may critique the logical relationship between program objectives and the activities designed to bring them about. Design evaluation, then, refers to the construct and logical or operational validity of a set of intentions. The standards involved in this sort of evaluation may not be made entirely explicit in advance or may be made explicit incrementally. The method used is that of logical argument and the evaluation itself is readily understood in terms of the SPD concepts.

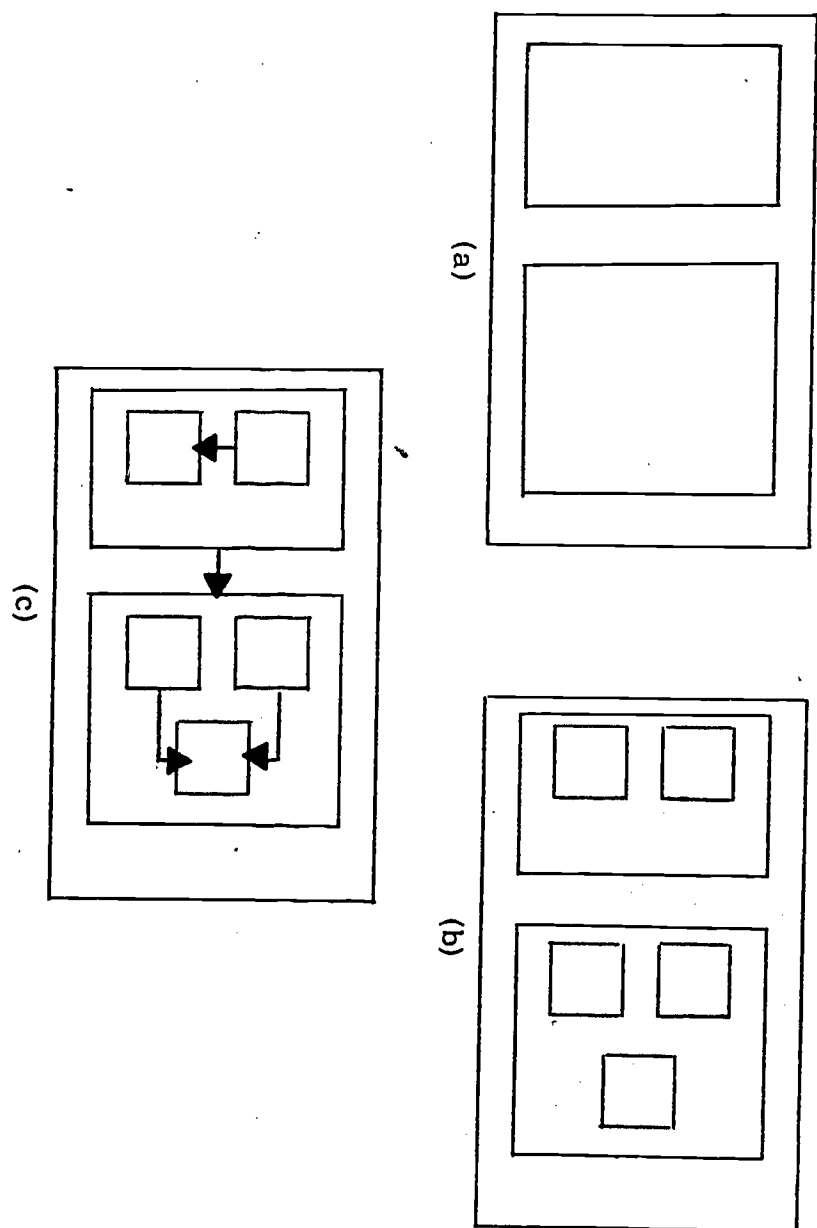
Program plans may themselves serve as Standard for other evaluation undertaken during the life cycle of the program. Program plans specify and direct program implementation and as such may serve as Standard to input, process, and outcome evaluation. For example, any program utilizes certain kinds and amounts of resources over time. Program plans which specify the number and kind of resources needed at different points in time relative to different activities and purposes can act as Standard governing the installation of the program. Thus, Performance information may be gathered concerning the extent to which the resources planned for are indeed available and are in fact deployed as required (input evaluation). This kind of information is exceptionally important in the management of a reasonably complex program. A program whose design has been judged adequate may nevertheless falter if it does not have the proper resources available when they are needed.

Process evaluation involves determining whether planned activities are carried out in the manner called for by the program plans and whether

they are of the quality expected. Again, the Standard here is the program plan which specifies and describes the program processes to be set into motion. Because of the complex interaction between Standard and action (which was known first?) thorough process evaluation overlaps with action research.

Outcome evaluation refers to determining the extent to which planned outcomes are achieved. It is useful to distinguish at least two classes of outcomes. Enabling or interim outcomes refer to milestones or subobjectives essential to the execution of the program from month to month. In contrast, terminal objectives refer to the major purposes or aims of the program. Again, a program plan can act as Standard relative to interim and terminal objectives to the extent that their interrelationships are known and explicit and indices of attainment have been specified.

In order to do design, input, process, and output evaluation and thus attempt to serve the program improvement interests and needs of program staff, a program must be planned in quite some detail. The plans must specify all major program activities as well as their resources and objectives. Figure 2 represents three levels of progressively deeper analysis into program operations which can make some of this detail available. The large rectangle in figure 2 (a) represents a program and the two rectangles inside it represent its two major components. In other words, the two smaller rectangles define the larger rectangle which contains them. Given any program, one may begin to produce the detailed plans necessary by initiating such a functional analysis and identifying major program components. A further analysis may be carried out simultaneously



Component Analysis of a Program

Figure 2

to complement the functional analysis by making use of three categories of information in terms of which an activity may be understood. An activity or component may be thought of as directed toward some end or else having a purpose. It may therefore be said to have some outcomes. There are also some things that are done in order to realize or bring about those outcomes, and these may be called processes. And finally, there are always some resources that are needed in order to support the processes, and these may be called inputs. Thus, for any program once its major components are identified, an input-process-output analysis may be performed for each.

The major components of a program may be further analyzed into subcomponents as shown in figure 2 (b). Input-process-output analysis may also be prepared for each of those subcomponents. In addition, by specifying the input-process-output dependencies among components and subcomponents, the relationship between all of those activities may be specified. These relationships are shown in figure 2 (c). Arrows show components or subcomponents having outputs which are inputs to other components or subcomponents. Each of the subcomponents in figure 2 (c) may, of course, in turn be analyzed into subcomponents and input-process-output descriptions may be prepared for each of them. This kind of analysis may be continued endlessly or as far as seems useful to program management.

It should be clear that properly specifying inputs, processes, and outputs for each component and subcomponent amounts to making the Standards essential to input process, and outcome evaluation available.

This makes it possible to conduct evaluation useful to program staff

because Performance data can be gathered relative to a larger class of program characteristics than just terminal objectives. The discrepancy information produced in the course of evaluation may be used to support two broad categories of management action. It may be used, on the one hand, to exert greater control over program operations in order to insure that Performance meets Standards. On the other hand, management may decide that the Standards originally set are inappropriate or unrealistic and may thus change the Standards involved.

We have seen that preparing for program evaluation can have the benefit of being identical with management-oriented program planning. In this way program evaluation may be conducted to satisfy those who are mostly interested in terminal outcomes but also to support some of the task-related judgments and decisions that management has to make on a continuous basis.

3. Profound standards. In section 2 immediately above, I outlined a way in which the programmatic aspects of a program can be analyzed and subjected to evaluation. Because this amounts to establishing an internal feedback system governing the major tasks of the program, the evaluation can serve program management needs directly. However, a program is usually also an organization. And there are dimensions of an organization other than its primary task structure which are equally as important to its satisfactory functioning and to attempts at program improvement. My experience leads me to suggest that the usefulness to program staff of program evaluation is related to the multiplicity of

dimensions looked at and the depth to which each dimension is considered. Therefore, in this section I would like to point to some dimensions beyond the task or organization subsystem of a program already considered which can also be subjected to the SPD approach outlined. The assumption continues to be that program improvement efforts require coordinated attention to the dynamics of all aspects of organizational behavior and functioning. A change in one dimension or aspect of an organization influences all its other aspects. Thus, evaluation aimed at program improvement must reach deeper into the organization than its task structure to exhaustively cover all dimensions of organizational functioning in the detail necessary to uncover and develop their relationship to the primary task of the program. However, a corollary is that in order to work with sufficient dimensions as needed, program staff must set and be the source of the Standards against which Performance will be compared. This requires the evaluator to distinguish carefully between his responsibilities and those of program management. Therefore, I first want to say something about the role of the evaluator as that person has been portrayed here.

a. The role of the Program Improvement Specialist (PIC).

In order to carry out the approach outlined in section 2 above, the evaluator must represent the programmatic decisions and procedures of management as faithfully as possible no matter how much he personally disagrees with them.² In other words, it must definitely be program management which must decide what the program Standards will be. The evaluator will help express and explicate them and confront management with decisions they

2. Obviously this assumes an accepted and shared moral and ethical structure underlying the program.

must make but ultimately the Standards against which Performance will be compared must be entirely those of program staff. While the evaluator will collect Performance information, it is up to management to identify the Performance information to be collected. That includes management having control over defining Performance indices and criterion Performance. Finally, the specific implications of the Performance information collected by the evaluator are also derived by program management. Management must decide what to do with the discrepancy information: whether to change the Standard involved or exert greater control over program operations so that discrepancy is reduced. Aggregating the discrepancy information into a judgment of worth about the particular aspect of the program being evaluated and making that judgment is also something that is done by program management. In that sense it is management that does the evaluating and not the evaluator.

This makes it confusing to keep on calling an evaluator whose aim is to directly assist program improvement efforts an evaluator. It might make more sense to call that person a program improvement consultant (PIC) or something similar. For we have seen that what a PIC does is facilitate a process which leads to and then supports evaluation as a line management function.

There is another way of looking at what the PIC does that helps to show how the application of the SPD concepts can be generalized to all aspects of organizational functioning. Since the SPD concepts can be used to describe much of individual human activity, we can say that evaluation as a line management function represents the application of consciously

articulated SPD cycles and thus the formalization of a very natural human activity. For example, in walking across a room, one chooses a path which acts as Standard, and as one walks, one continually compares Performance with Standard in order to stay on course. If for some reason, one wanders off course as judged by a comparison between Performance and Standard, one can take corrective action in order to get back on course. It may also happen, of course, that one encounters some obstacles which at first seemed manageable, but now, on trial, turn out to be insurmountable. In that case, a new route has to be discerned, i. e., the Standard has to be changed before continuing. Similarly, the role of expectations and purposes in human behavior can be analyzed in terms of the SPD concept as can any human action that is based on a recognized or incompletely recognized assessment of consequences. Again, it may be that the Standards governing action may be hard to specify or perhaps in important respects simply unrecognized, but in principle, one may speak of SPD when describing human actions. Thus, if we say that SPD cycles lie at the bottom of most human action, then formal internal evaluation is represented by those SPD cycles governing human action which are made explicit and are communicated to others. In other words, the PIC is concerned with making SPD cycles relative to all classes of activity explicit and visible in cooperation with those with whom he works. This implies commitment to a value system which favors supporting decision making with empirical data and which shows an emerging or substantial preference for consciously articulating basic assumptions and basic ingredients of decision making.

The additional dimensions made available in section D2 immediately above, beyond just determining whether terminal objectives were accomplished, are a result of the application of certain concepts. That is, the additional aspects of program operation which were explicated and for which Standards can be articulated are a direct function of the application of a model (component and input-process-output analysis) to program design and operation. What we need now are concepts that permit us to model an organization as a living whole. In what follows, I will outline some concepts or approaches that help map an organization comprehensively in all its dimensions.* The implication is that the PIC must be able to work in all these dimensions with the data peculiar to each and therein lie the implications for training PICs. We will see that we can think of the PIC as an organizational therapist and an applied social and moral philosopher.

b. Sociotechnical systems. It is useful to begin by positing that educational evaluation aimed at program improvement cannot restrict itself to examining what the schools do, pure and simple. There are a large number of other institutions and organizations, such as the family, organized religion, the advertising industry, etc., which act as shaping forces. These both embody and promote values and ideas about what is good and desirable and they seek to change behavior and influence thinking and attitudes accordingly. Moreover, some of these bodies may well exert a more consistently powerful influence than do the schools themselves. For example, the advertising industry may be considered to implicitly be promoting dependency and passiveness. Overeating is legitimized and subtly encouraged--its immediate consequences being relieved through a

*This will be done in very abbreviated fashion here but enough, I hope, to point out a direction.

of harness run and not effectively challenged for the way of life it represents and creates. Similarly, successful business practice is most often measured by number of sales, income, or size, to the point where the end becomes not the production of essential items or the delivery of services but making money and gaining power.

In contrast to tribal societies or non-industrialized cultures, life in the United States has come to be characterized by cultural flux, ambiguity and uncertainty. More often than not it is not obvious what the right or acceptable thing is to do, what career should be attempted, in brief, what one's social niche is when the context is changing in unpredictable ways and niches need to be made or consciously selected rather than accepted uncritically. In the face of all this the school can only consider itself an entity with an ambiguous charter. What does seem clear is that the traditional notion of transmitting reconstructed subject matter knowledge through didactic instruction in a narrow range of arbitrary disciplines is an insufficient preoccupation for a school. If we grant that other interests and groupings in society also influence values, attitudes, behaviors, and

beliefs, and if these conflict, which influences should be considered educational, that is, desirable? The answer is not likely to rest with identifying some that are and some that aren't but in a continuous examination of what the influences and their consequences are. The practical meaning of this for the schools, as educational agencies, is that they must take the initiative in defining their own role. And they must do that in relation to all other influences they can discern. In other words, schools must set their own context-specific goals and define their place and role in the community they serve based on continuous examination of events around them.

Figure 3 shows that the function of education is being undertaken by many different interests, groups, or institutions, the school being just one of them. Designing a school program thus involves reaching some understanding.

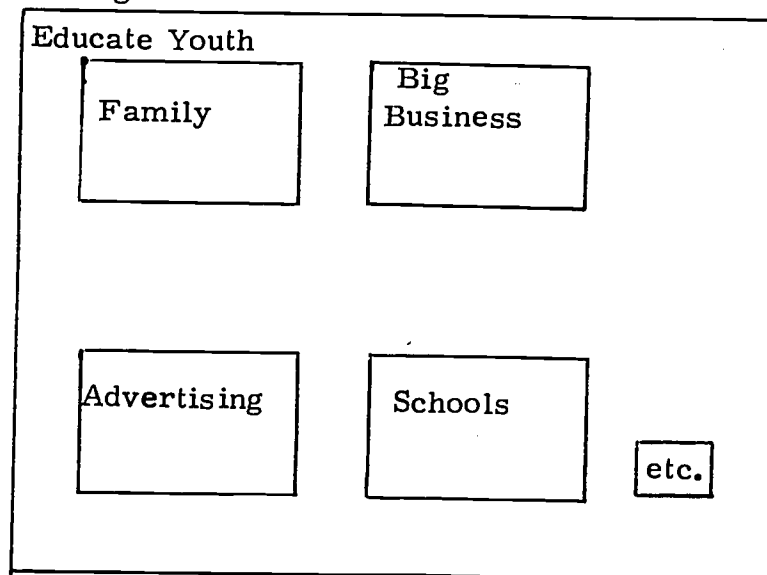


Figure 3:
Education in Society

(or forming some belief) about the influences exerted by these other agencies and about their interdependencies. That is not easy since their avowed goals are not necessarily their goals in retrospect and their actual effects are not always discernible nor predictable. But without such a continuous analysis the school runs an intolerably large risk of being irrelevant or unrealistic. To carry out its educative role the school must identify forces (ideas, pressures, etc.) it chooses to oppose or counteract and forces it chooses to supplement, revitalize, or initiate. In that respect a school program would always in part be engaged in bringing latent social issues into consciousness. As I am portraying it here, it would be the PIC who catalyzes this analysis, and that makes the PIC part applied social and moral philosopher. (What will emerge too is that the functions of the PIC are best carried out by a team of people working together.)

Beyond taking into account forces in the environment which bear upon the definition of a school's purposes and programs, a school must take into account forces and phenomena internal to it as an organization, in so far as these also bear upon its purposes and activities. Thus, we will now turn our attention from the problems surrounding the definition of the external boundary of the school to the dynamics of the school as an organization.³

To think about a school only in terms of its separate programs is to miss taking into account much of the "treatment." A particular program may design specific experiences aimed at specific ends but it also must judiciously exhibit the learning involved in the design of the

3. It should be clear that the discussion applies as well if we are considering a program within a school.

very same experiences. It must exercise the learnings which are the topics of the different programs in the school at every pertinent opportunity. This might as well be done deliberately since it can safely be assumed that the social system that is the school will exert its own influence on all of the organized activities which the school undertakes. Unless these influences are shaped, directed and managed, they will grow out of recognition and become part of the hidden curriculum--which eventually leads to the pseudo-puzzle of "the more things change the more they stay the same."

What is needed therefore is a practical way of thinking of the school (or a program) as an organization--a way which permits the members of the organization to identify and direct the various ways they have open to design an educational environment. One way of doing this is to think of the system or organization involved as a sociotechnical system. This means describing the system using three dimensions: the technical, the structural, and the psychosocial.⁴ The technical dimension refers to the technology of an organization, to its tools, concepts, techniques, facilities, equipment and so on; the psychosocial dimension refers to the "human" side of the organization--to norms, values, needs, desires, and generally to intrapsychic and interpersonal factors; and finally structure refers to organizing human activities around the organization's technology and psychosocial dimension. The structure of an organization mediates between its technical and psychosocial aspects. A proper structure is sensitive and responsive to both instead of to just one as so often happens when a structure is oriented exclusively to the convenience of the task or production.

4. This way of talking about sociotechnical systems comes from a presentation on sociotechnical systems given by Thomas Dolgoff, The Menninger Foundation, December 1974. The concept of sociotechnical systems is developed in A. K. Rice, The Enterprise and Its Environment, and elsewhere.

What we can say now of the approach described in section D2 above is that it tends to be limited to the relationship between technology and structure. The component and input-process-output analysis of the primary task of an organization helps define what knowledges or attitudes are to be made the subject of specific instruction, how resources will be deployed, how personnel will be assigned, what record keeping and management procedures are appropriate to the task, and so on. What it does not examine is how the program's technology relates to the psychosocial dimension nor what kinds of needs, wants, norms, etc., should be given expression over others and in what circumstances. Thus, Performance Standards are set relative to task execution leaving the interaction between structure and psychosocial factors and technology and psychosocial factors largely implicit or unexamined. Yet a program can fail as easily because the right number of teachers were not available as it can because, even though they are available, an informal structure is established based on unrecognized emotional coping mechanisms which work to the disadvantage of all.

c. Training implications for PICs. Looking at an organization in terms of its technology, structure, and psychosocial factors enlarges the PIC's sphere of activity. It certainly requires different skills to articulate and clarify standards for the technical as opposed to the psychosocial part of an organization and to facilitate the design of a structure sensitive to both, not just one. There are two major training implications for a P which result from this way of looking at things which I would like to briefly consider here through some examples.

The first one is the ability to build models. This refers to the ability to identify and order as many different variables and factors as possible that have a bearing on a situation in order to represent its dynamics in a realistic and coherent way. It includes knowledge of available representations of how organizations work and the ability to apply them and draw from them as needed. It also refers to facility with metamodels or with concepts and procedures (such as input, process, output analysis) which can be used in order to generate models. For example, being able to draw up a network showing major program components and create input, process, output descriptions for each through analysis of documents and interviews represents an application of meta-concepts and results in a model of a program (which can act as standard). A PIC must have the ability to map or model all dimensions of a program and to show external and internal forces and interfaces.

The second ability is to be able to gather data relative to a large variety of different phenomena and to report this data using print, verbal, nonverbal, visual, and auditory media. If, for example, the standard is that all students should be able to multiply real numbers, then the PIC might use standard testing techniques in order to gather the data needed to see whether what is matches what should be (i. e., he takes Performance measures and generates discrepancy information by comparing Performance to Standard). Methods in statistics, measurement, and experimental design would be appropriate here. But if the Standard is that teachers be able to express to each other the ways in which they consider each other's contributions at curriculum planning meetings helpful and not helpful in accomplishing their tasks, then the PIC must be able to know

how to observe interaction between people and gather specific performance information that shows the ways in which that Standard is and is not being met. And different data is analyzed and reported differently. A written report on test results one week after testing may be the appropriate thing to do in the first example above. Program management can use such a report in judging the adequacy of instruction or other conditions related to the math achievement of the students involved. But in the second example, keeping in mind the purpose of program improvement, it is likely that to be effective the PIC report must be verbal and person and situation specific. The PIC might attend some of the teacher's curriculum planning meetings and intervene directly during the course of the meeting or immediately after the meeting. The PIC might, in one move, report Performance information and contrast it to the Standard in order to make it sensible. This would focus attention on both the Standard and Performance at one moment with regard to a very specific situation and thus require reaffirmation of the Standard in a way which is not necessary to do in the case of the Standard governing mathematics achievement in the example cited above. What the PIC could easily be doing here is to help teachers clarify the Standard, understand how it pertains to a specific behavior of theirs, redefine the Standard as it applies to them if necessary, or formulate action specific to themselves as individuals and to the group as a whole in order to insure compliance with the Standard.

The skills called for on the part of the PIC in setting Standards in both of these examples are significantly different and this returns us to the notion of model building already mentioned. A component and input-

process-output analysis led by the PIC may be sufficient to identify and define the objective of having all students be able to multiply real numbers and be sufficient to set a usable and realistic Standard. But the Standard having to do with the ability of teachers to share their feelings relative to their tasks requires a significantly different approach and set of skills. The problem as originally presented to the PIC may consist of a request that the curriculum development component of the program be "evaluated." More specifically, management may be aware that that component isn't functioning properly, but may need help in specifying exactly how it should be functioning, i. e., in specifying what should be done by whom and to what end. In other words, management needs a Standard for the curriculum development component against which it may compare specific Performance and so judge the adequacy of the component and take action as needed. In helping to set the necessary Standards (such as, in the example given here, that teachers should share their impressions of each other's contributions), the PIC would find it useful to be familiar with models governing human interaction and effective teamwork. Facility in that arena is as helpful to management in that situation as is doing an input-process-output analysis of math instruction in a different one.

d. Generic subsystem categories. Existing models of organizations provide additional ways of describing the approach a PIC might take. For example, Katz and Kahn (1966)⁵ describe an organization as consisting of a set of related subsystems each making a very specific and important contribution to the whole (see Figure 4).

5. Katz and Kahn, The Social Psychology of Organizations, John Wiley, 1966.

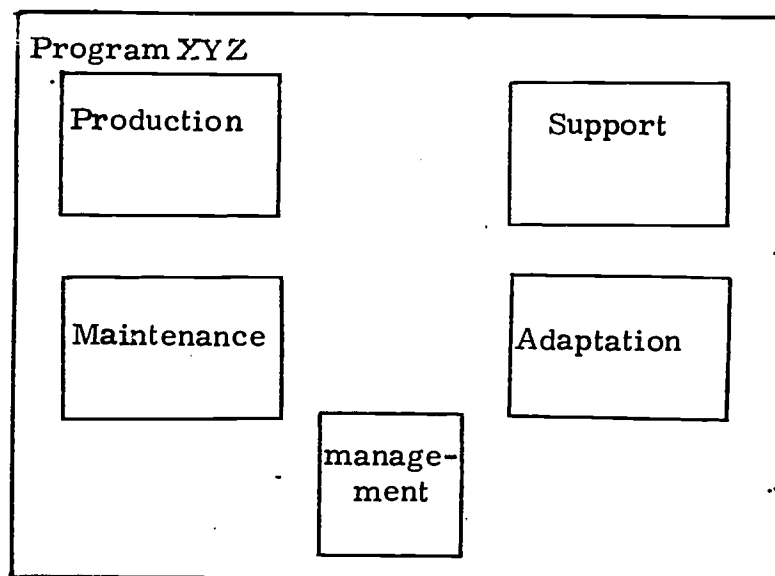


Figure 4: Generic Subsystem Categories Encompassing Any Organization

The production subsystem refers to the activities which directly comprise the primary task of the organization. Thus, in a school, classroom instruction and other teaching activities would be subsumed by this subsystem. The support subsystem deals with procuring input and dispersing output and is thus closely tied to the production subsystem. It is also helpful to include here the liaison activities undertaken by any organization in order to bring about satisfactory relations between the organization and the environment. The maintenance subsystem refers to the activities directed at maintaining the equipment, materials, people, etc., that get the primary task of the organization done. It includes recruitment, reward, and orientation functions and is geared to dealing with morale problems and with tying people into the proper activities in the system. Thus, this subsystem is concerned with keeping both machines and people properly lubricated and in good physical and spiritual conditions. The adaptive subsystem is concerned with the survival of the organization in a changing environment. It includes research and development, planning, and marketing activity and the general function of scanning the environment for trends or

opportunities important to the livelihood of the total organization. And finally, there is the management subsystem which directs and controls the organization as a whole. It includes the regulatory function (for example, internal evaluation) and the authority structure which allocates decision making.

With this general model as a guide, the PIC can help set up SPD cycles relative to any given aspect of an organization which are sensitive to the interpenetration of the different generic subsystems and the bearing they have on a particular problem. By thinking in terms of these generic categories, and the functions, contributions, and roles they summarize, the PIC can increase the dimensions of organizational functioning treated in any given SPD cycle, and thus directly assist the program improvement efforts of program staff. This approach to program evaluation aimed at program improvement reaches even deeper into the program to encompass all possible dimensions in its SPD cycles.

e. Parson's theory of action. One other conceptual framework or model is worth mentioning here for the power it can give the work of a PIC. It is the theory of action developed by Parsons⁶ and it complements the sociotechnical concepts and generic subsystem categories already outlined by providing a model governing organized endeavor. Parsons lays out two fundamental dilemmas faced by organized systems: first, since every system functions in relation to an environment, there is the problem of choosing between working to integrate the different units within the organization

6. See for example Talcott Parsons, The Social System, Free Press of Glencoe, 1951.

and working to achieve the most advantageous or beneficial relationship with the environment. Both must always be attended to over time. And second, there is the problem of choosing between stability over time or delayed gratification, and consummation or immediate gratification. These two dilemmas lead to what he calls four functional imperatives:

- (a) adaptation: stability over time in relation to the environment
- (b) pattern-maintenance: stability over time in relations among units
- (c) goal attainment: gratification in relation to environment
- (d) integration: gratification in relations among units.

These four fundamental subsystems help us see how goal seeking processes are guided by both the consideration of empirical means-ends relations and normative factors, something of great importance to evaluation efforts aimed at program improvement. They help see how action is an expression of means-ends considerations as much as it is an expression of meaning; how action can be understood in terms of a theory of object-relations in which action is shaped by aspects of the situation as much as by a shared system of values. This leads to the PIC as organizational therapist.

There is one last thing I would like to say about the relationship between program evaluation and program improvement. Inevitably, when doing program evaluation one is faced with the interest, usually on the part of funding agents, for "proven products," "generalizable packages," and so on. This leads to approaches to evaluation guided by the idea that what needs to be done is to find the method that works. The work of the PIC

as I have tried to portray it here, however, is guided by the idea that if we want to establish what it is that works then we have to understand that what we are doing is describing a rose or perhaps a violet. We are not establishing a prescription of what works for others to follow. Others have to discover that they are Daisy or Iris. For what we are describing are interactions between people and all interactions are different from each other in spite of the advertisers who tell us to do this to be masculine and that to be feminine and attractive. Advertising of that sort is destructive and furthers alienation because it is built on falseness. And scientific method harnessed to finding out the way of organizing social action will be equally destructive and alienating. In trying to be a rose when you are a violet you destroy yourself.